確率・統計基礎: 期待値、分散、モーメント

森 立平

確率空間と確率変数

確率空間 (Ω, P)

- $\Omega\colon$ (非可算でもよい) 集合 (標本空間)
- $P\colon 2^{\Omega} \to [0,1]$ (確率測度)

確率の公理

1. $P(\Omega) = 1$.
2. $\forall (A_n \subseteq \Omega)_{n=1,\dots}$,
   $\forall i \neq j, A_i \cap A_j = \varnothing \implies P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$
   (完全加法性, σ-加法性).

確率変数 (確率空間 (Ω, P) に付随する)
$X\colon \Omega \to \mathbb{R}$.

$$\Pr(X \in A) := P(X^{-1}(A)) \qquad \forall A \subseteq \mathbb{R}$$

確率変数から定まる確率空間

Example

- $\Omega = \{\mathrm{H}, \mathrm{T}\}^2 = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$.
- $P(\{\mathrm{HH}\}) = P(\{\mathrm{HT}\}) = P(\{\mathrm{TH}\}) = P(\{\mathrm{TT}\}) = 1/4$.
- $X(\mathrm{HH}) = 2,\ X(\mathrm{HT}) = X(\mathrm{TH}) = 1,\ X(\mathrm{TT}) = 0$.

このとき、

$$\begin{aligned}
\Pr(X = 0) &= P(\{\mathrm{TT}\}) = 1/4 \\
\Pr(X = 1) &= P(\{\mathrm{HT}, \mathrm{TH}\}) = 1/2 \\
\Pr(X = 2) &= P(\{\mathrm{HH}\}) = 1/4
\end{aligned}$$

である

そもそも最初から確率空間

- $\Omega = \{0, 1, 2\}$.
- $P(\{0\}) = 1/4, P(\{1\}) = 1/2, P(\{2\}) = 1/4$

を考えればよいのでは？

複数の確率変数

Example

- $\Omega = \{\mathrm{H}, \mathrm{T}\}^2 = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$.
- $P(\{\mathrm{HH}\}) = P(\{\mathrm{HT}\}) = P(\{\mathrm{TH}\}) = P(\{\mathrm{TT}\}) = 1/4$.
- $X(\mathrm{HH}) = 2,\ X(\mathrm{HT}) = X(\mathrm{TH}) = 1,\ X(\mathrm{TT}) = 0$.
- $Y(\mathrm{HH}) = Y(\mathrm{HT}) = 1,\ Y(\mathrm{TH}) = Y(\mathrm{TT}) = 0$.

このとき、

$$\begin{aligned}
\Pr(X = 0,\ Y = 0) &= P(X^{-1}(0) \cap Y^{-1}(0)) = P(\{\mathrm{TT}\}) = 1/4 \\
\Pr(X = 0,\ Y = 1) &= P(X^{-1}(0) \cap Y^{-1}(1)) = P(\varnothing) = 0 \\
\Pr(X = 1,\ Y = 0) &= P(X^{-1}(1) \cap Y^{-1}(0)) = P(\{\mathrm{TH}\}) = 1/4 \\
\Pr(X = 1,\ Y = 1) &= P(X^{-1}(1) \cap Y^{-1}(1)) = P(\{\mathrm{HT}\}) = 1/4 \\
\Pr(X = 2,\ Y = 0) &= P(X^{-1}(2) \cap Y^{-1}(0)) = P(\varnothing) = 0 \\
\Pr(X = 2,\ Y = 1) &= P(X^{-1}(2) \cap Y^{-1}(1)) = P(\{\mathrm{HH}\}) = 1/4
\end{aligned}$$

である。複数の確率変数を考えるときには確率空間は必要。

連続確率空間

$\Omega = \mathbb{R}$ の場合の確率空間 $(\mathbb{R}, P)$ について考える。

ある積分できる関数 $p\colon \mathbb{R} \to \mathbb{R}$ が存在して

$$P([a, b]) = \int_a^b p(x)\,\mathrm{d}x$$

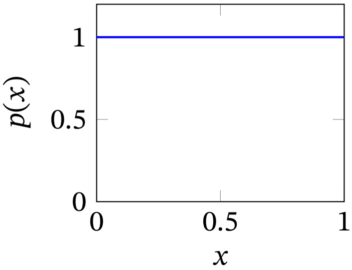
であるとき、p を確率密度関数という。

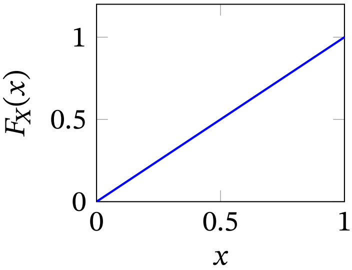
確率密度関数は常に存在するとは限らないが累積分布関数
$F_X(a) = \Pr(X \leq a)$ は常に存在する。

一様分布

$$\Omega = [0, 1]$$

$$p(x) = 1, \qquad \Pr(X \le x) = x$$

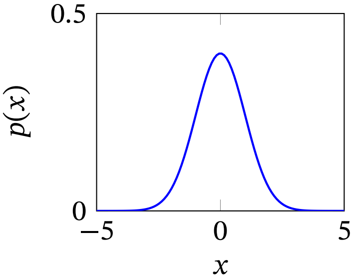


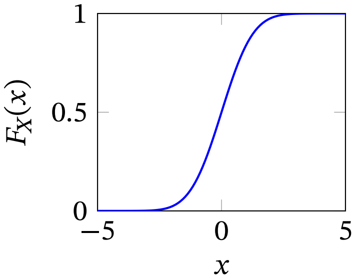


標準正規分布 (標準ガウス分布)

$$\Omega = \mathbb{R}$$

$$p(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}}, \quad \Pr(X \leq x) = \Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}} \, \mathrm{d}t$$

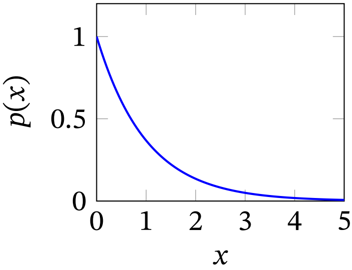


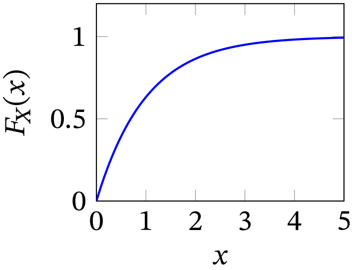


指数分布

$$\Omega = [0,\ \infty)$$

$$p(x) = \lambda e^{-\lambda x}, \qquad \Pr(X \le x) = 1 - e^{-\lambda x}$$

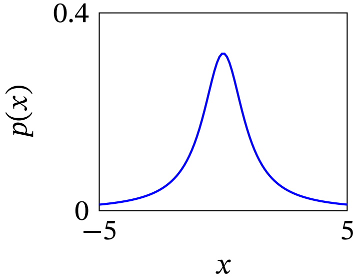


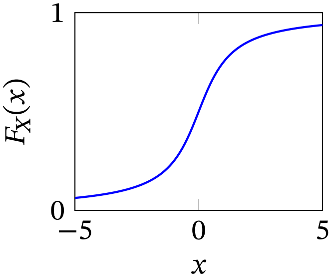


標準コーシー分布

$$\Omega = \mathbb{R}$$

$$p(x) = \frac{1}{\pi(1 + x^2)}, \qquad \Pr(X \leq x) = \frac{1}{\pi} \arctan(x) + \frac{1}{2}$$





期待値 (平均)

非負離散確率変数 X について期待値を以下で定義する。

$$\mathbb{E}[X] := \sum_k \Pr(X = k)k$$

離散確率変数 X について $\mathbb{E}[|X|] = \sum_k \Pr(X = k)|k|$ が存在するとき、

$$\mathbb{E}[X] := \sum_k \Pr(X = k)k$$

連続確率変数 X について

$$\mathbb{E}[X] := \int p(x)x \, \mathrm{d}x$$

(上記の積分が広義リーマン積分のときは条件に $\int p(x)|x| \, \mathrm{d}x < \infty$ が必要)

期待値の線形性

絶対収束するので和の順番は自由に変えてよい

$$
\begin{aligned}
\mathbb{E}[f(X)] &= \sum_k \Pr(f(X) = k)k \\
&= \sum_k \Pr(X \in f^{-1}(k))k \\
&= \sum_k \sum_{t \in f^{-1}(k)} \Pr(X = t)k \\
&= \sum_t \Pr(X = t)f(t)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}[aX + bY] &= \sum_k \Pr(aX + bY = k)k \\
&= \sum_{x,y} \Pr(X = x, Y = y)(ax + by) \\
&= \sum_{x,y} \Pr(X = x, Y = y)ax + \sum_{x,y} \Pr(X = x, Y = y)by \\
&= \sum_x \Pr(X = x)ax + \sum_y \Pr(Y = y)by \\
&= a\sum_x \Pr(X = x)x + b\sum_y \Pr(Y = y)y = a\mathbb{E}[X] + b\mathbb{E}[Y]
\end{aligned}
$$

積の期待値

独立な確率変数 X と Y について

$$
\begin{aligned}
\mathbb{E}[XY] &= \sum_k \Pr(XY = k)\,k \\
&= \sum_{x,y} \Pr(X = x, Y = y)\,xy \\
&= \sum_{x,y} \Pr(X = x)\Pr(Y = y)\,xy \\
&= \left(\sum_x \Pr(X = x)x\right)\left(\sum_y \Pr(Y = y)y\right) \\
&= \mathbb{E}[X]\,\mathbb{E}[Y]
\end{aligned}
$$

期待値の例

Example (ベルヌーイ分布)

$$\Pr(X = 0) = 1 - p$$
$$\Pr(X = 1) = p$$

のとき、

$$\begin{aligned}\mathbb{E}[X] &= \Pr(X = 0) \cdot 0 + \Pr(X = 1) \cdot 1 \\ &= p\end{aligned}$$

期待値の例

Example (二項分布)

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

のとき、

$$\begin{aligned}\mathbb{E}[X] &= \sum_k \Pr(X = k)k \\ &= \binom{n}{k} p^k (1-p)^{n-k} k\end{aligned}$$

一方で t 回目のコイン投げで表が出たら 1、裏が出たら 0 という確率変数を X_t とおくと、$X = X_1 + X_2 + \cdots + X_n$ なので

$$\mathbb{E}[X] = \mathbb{E}[X_1 + \cdots + X_n] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] = np.$$

期待値の例

Example (標準正規分布)

$$p(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}}$$

のとき、

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} p(x) x \, \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} x \, \mathrm{d}x \\
&= \left[-\frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} \right]_{-\infty}^{\infty} = 0
\end{aligned}$$

期待値の例

Example (正規分布)

$$p(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(x-m)^2}{2}}$$

のとき、

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} p(x) x \, \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(x-m)^2}{2}} x \, \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} (x+m) \, \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} m \, \mathrm{d}x = m
\end{aligned}$$

期待値の例

Example (コーシー分布)

$$p(x) = \frac{1}{\pi(1+x^2)}$$

のとき、

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} p(x) x \, \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{1}{\pi(1+x^2)} x \, \mathrm{d}x \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\frac{1}{x+i} + \frac{1}{x-i} \right) \mathrm{d}x
\end{aligned}$$

マルコフの不等式

Theorem (マルコフの不等式)

任意の非負の確率変数 X と $a > 0$ について

$$\Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

Proof.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k} \Pr(X = k)k \\
&= \sum_{k \geq a} \Pr(X = k)k + \sum_{k < a} \Pr(X = k)k \\
&\geq \sum_{k \geq a} \Pr(X = k)k && (k \geq 0) \\
&\geq \sum_{k \geq a} \Pr(X = k)a \\
&= a \Pr(X \geq a).
\end{aligned}$$

□

分散

$$
\begin{aligned}
\mathbb{V}[X] &:= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \\
&= \mathbb{E}\left[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2\right] \\
&= \mathbb{E}\left[X^2\right] - 2\mathbb{E}[X]\,\mathbb{E}[X] + \mathbb{E}[X]^2 \\
&= \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2
\end{aligned}
$$

分散は非負で期待値からの広がりを表す量。

$$
\mathbb{V}[X] = 0 \iff \Pr(X = \mathbb{E}[X]) = 1.
$$

分散の性質

$$\mathbb{V}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

$$\begin{aligned}
\mathbb{V}[aX] &= \mathbb{E}\left[(aX - \mathbb{E}[aX])^2\right] \\
&= \mathbb{E}\left[(aX - a\mathbb{E}[X])^2\right] \\
&= \mathbb{E}\left[a^2(X - \mathbb{E}[X])^2\right] \\
&= a^2\mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \\
&= a^2\mathbb{V}[X]
\end{aligned}$$

$$\begin{aligned}
\mathbb{V}[X+Y] &= \mathbb{E}\left[(X + Y - \mathbb{E}[X+Y])^2\right] \\
&= \mathbb{E}\left[(X - \mathbb{E}[X] + Y - \mathbb{E}[Y])^2\right] \\
&= \mathbb{E}\left[(X - \mathbb{E}[X])^2 + (Y - \mathbb{E}[Y])^2 + 2(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right] \\
&= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] + \mathbb{E}\left[(Y - \mathbb{E}[Y])^2\right] + 2\mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right] \\
&= \mathbb{V}[X] + \mathbb{V}[Y] + 2\color{red}{\mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right]}
\end{aligned}$$

共分散

$$\begin{aligned}\mathrm{Cov}[X, Y] &:= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]\end{aligned}$$

- $X - \mathbb{E}[X]$ と $Y - \mathbb{E}[Y]$ の符号が一緒 $\Rightarrow \mathrm{Cov}[X, Y] \geq 0$。
- $X - \mathbb{E}[X]$ と $Y - \mathbb{E}[Y]$ の符号が逆 $\Rightarrow \mathrm{Cov}[X, Y] \leq 0$。

直感的な意味

- X が大きいとき Y が大きい $\Rightarrow \mathrm{Cov}[X, Y] \geq 0$。
- X が大きいとき Y が小さい $\Rightarrow \mathrm{Cov}[X, Y] \leq 0$。

無相関と独立

$$\mathrm{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

確率変数 X と Y が無相関

$$\stackrel{\text{def}}{\iff} \mathrm{Cov}[X, Y] = 0 \iff \mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$$

確率変数 X と Y が独立 $\Longrightarrow \mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y] \iff$ 確率変数 X と Y が無相関

無相関だけど独立でない例

$$\Pr(X = 1, Y = 0) = \frac{1}{2}$$
$$\Pr(X = -1, Y = 1) = \frac{1}{4}$$
$$\Pr(X = -1, Y = -1) = \frac{1}{4}$$

n 個の確率変数の和の分散

$$
\begin{aligned}
\mathbb{V}\left[\sum_{k=1}^{n} X_k\right] &= \mathbb{E}\left[\left(\sum_{k=1}^{n} X_k - \mathbb{E}\left[\sum_{k=1}^{n} X_k\right]\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_{k=1}^{n} (X_k - \mathbb{E}\left[X_k\right])\right)^2\right] \\
&= \mathbb{E}\left[\sum_{k=1}^{n} (X_k - \mathbb{E}\left[X_k\right])^2 + 2 \sum_{1 \le k < \ell \le n} (X_k - \mathbb{E}\left[X_k\right])(X_\ell - \mathbb{E}\left[X_\ell\right])\right] \\
&= \sum_{k=1}^{n} \mathbb{V}\left[X_k\right] + 2 \sum_{1 \le k < \ell \le n} \mathrm{Cov}\left[X_k, X_\ell\right]
\end{aligned}
$$

任意の $1 \le k < \ell \le n$ について X_k と X_ℓ が独立
$\implies \mathrm{Cov}\left[X_k, X_\ell\right] = 0, \forall k < \ell$

$X_1, \dots, X_n$ がペア独立 (pairwise independent) $\overset{\text{def}}{\iff}$ 任意の $1 \le k < \ell \le n$ について X_k と X_ℓ が独立

チェビシェフの不等式

Lemma (チェビシェフの不等式)

確率変数 X が分散を持つと仮定する。任意の $a > 0$ について

$$\Pr(|X - \mathbb{E}[X]| \geq a) \leq \frac{\mathbb{V}[X]}{a^2}$$

Proof.

$$\begin{aligned}
\Pr(|X - \mathbb{E}[X]| \geq a) &= \Pr((X - \mathbb{E}[X])^2 \geq a^2) \\
&\leq \frac{\mathbb{E}\left[(X - \mathbb{E}[X])^2\right]}{a^2} \qquad \text{(マルコフの不等式)}
\end{aligned}$$

□

モーメント母関数

確率変数 X の n 次モーメント: $\mathbb{E}[X^n]$.

確率変数 X のモーメント母関数 (積率母関数): $M_X(t) := \mathbb{E}\left[e^{tX}\right]$.
モーメント母関数が $t \in (-\epsilon, \epsilon)$ で存在するとき、この範囲で

$$\begin{aligned} M_X(t) &= \mathbb{E}\left[e^{tX}\right] \\ &= \mathbb{E}\left[\sum_{n \geq 0} \frac{(tX)^n}{n!}\right] \\ &= \sum_{n \geq 0} \mathbb{E}\left[\frac{(tX)^n}{n!}\right] \qquad \text{(極限と積分の交換)} \\ &= \sum_{n \geq 0} \frac{\mathbb{E}[X^n]}{n!} t^n \\ &= 1 + \mathbb{E}[X]\, t + \frac{\mathbb{E}[X^2]}{2} t^2 + \cdots \end{aligned}$$

$$\left.\frac{d^n M_X(t)}{dt^n}\right|_{t=0} = \mathbb{E}[X^n] \qquad n \geq 0$$

モーメント母関数の例

Example (二項分布)

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

のとき、

$$\begin{aligned} M_X(t) &= \mathbb{E}[\mathrm{e}^{tX}] \\ &= \sum_k \Pr(X = k)\mathrm{e}^{tk} \\ &= \sum_k \binom{n}{k} p^k (1-p)^{n-k} \mathrm{e}^{tk} \\ &= (p\mathrm{e}^t + (1-p))^n \end{aligned}$$

$$\left.\frac{\mathrm{d}M_X(t)}{\mathrm{d}t}\right|_{t=0} = n(p\mathrm{e}^t + (1-p))^{n-1} p\mathrm{e}^t\Big|_{t=0} = np$$

独立確率変数の和のモーメント母関数

独立確率変数 X と Y について

$$
\begin{aligned}
M_{X+Y}(t) &= \mathbb{E}\left[\mathrm{e}^{t(X+Y)}\right] \\
&= \mathbb{E}\left[\mathrm{e}^{tX}\right]\mathbb{E}\left[\mathrm{e}^{tY}\right] \\
&= M_X(t)M_Y(t)
\end{aligned}
$$

ベルヌーイ分布のモーメント母関数は

$$
\begin{aligned}
M_X(t) &= \mathbb{E}\left[\mathrm{e}^{tX}\right] \\
&= p\mathrm{e}^t + (1-p)
\end{aligned}
$$

二項分布に従う確率変数はベルヌーイ分布に従う n 個の独立確率変数の和なので、そのモーメント母関数は

$$
(p\mathrm{e}^t + (1-p))^n
$$

キュムラント母関数

キュムラント母関数 $K_X(t)$ は以下で定義される。

$$K_X(t) := \log(M_X(t)) = \log \mathbb{E}\left[e^{tX}\right]$$

$$\left.\frac{\mathrm{d}K_X(t)}{\mathrm{d}t}\right|_{t=0} = \mathbb{E}[X], \quad \left.\frac{\mathrm{d}^2K_X(t)}{\mathrm{d}t^2}\right|_{t=0} = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = \mathbb{V}[X]$$

独立確率変数 X と Y について

$$K_{X+Y}(t) = K_X(t) + K_Y(t).$$

課題

- 標準正規分布のモーメント母関数、期待値、分散をもとめよ。
- $\lambda > 0$ について、非負整数上の分布であるポアソン分布 $\Pr(X = k) = \dfrac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$ のモーメント母関数、期待値、分散をもとめよ。
- $\lambda > 0$ について、非負実数上の分布である指数分布 $p(x) = \lambda \mathrm{e}^{-\lambda x}$ のモーメント母関数、期待値、分散をもとめよ。モーメント母関数が存在する t の範囲も示すこと。